THE USE OF KRIGING IN STOCHASTIC MODEL UPDATING AND ITS EFFECT ON PARAMETER ESTIMATES

Liang, P., DiazDelaO, F.A., Patelli, E., and Mottershead, J.E.

Institute of Risk and Uncertainty, School of Engineering, University of Liverpool, Liverpool L69 3GH, UK.
E-mail of the corresponding author: j.e.mottershead@liverpool.ac.uk

SUMMARY: The Kriging predictor or Gaussian process emulator is a commonly used surrogate model for reducing the computational cost in the forward propagation of model updating. It has already been used and discussed by many other researchers. As a stochastic model, the Kriging predictor provides not only the mean value of the prediction but also the prediction variance. Previous work mainly used the prediction variance as a kind of judgement on the model accuracy during the selection of design points and took only the mean value of the predictor in the further study and production of results. Therefore, the model updating result obtained from only the mean value of the Kriging predictor may be considered not accurate enough because of neglecting the prediction variance information. In order to investigate the effect of the Kriging prediction variance on parameter estimates, an approach is developed in this paper for including it in model updating procedure. The approach is based on a transfer method for the prediction variance and bootstrap algorithm. The main steps and the assumptions are presented, followed by numerical examples using a univariate model updating problem. The results obtained show the effect of the Kriging prediction variance on parameter estimates and are compared to results produced by conventional methods that neglect the Kriging prediction variance in model updating.

KEYWORDS: Model updating, Kriging, prediction variance.

1. INTRODUCTION

The Kriging model is a very commonly used interpolation technique. It provides an efficient surrogate model for the expensive forward propagation in model updating. Previous work on model updating with the Kriging model mainly considered the mean value of the Kriging prediction. However, the prediction variance of the Kriging model also affects the model-updating result. It indicates the accuracy of the Kriging prediction and therefore also affects the parameters estimated by the model updating process using the Kriging model. The Kriging prediction variance was used for selecting new design points to obtain a better Kriging model [1][2][3], for global optimisation of time-consuming Monte-Carlo simulation [4], and for measuring the quality of the prediction of a Kriging model [5]. In this paper, the Kriging prediction variance will be used with a variance transfer method and the bootstrap technique in order to estimate the its effect on the parameters estimated by model updating with Kriging.
The rest of this article is organised as follows. The univariate problem concerned will be presented in Section 2. Section 3 introduces the background theory of model updating using the Kriging model and kernel density estimation. Section 4 explains the procedure for including the Kriging prediction variance in the model updating result. The numerical examples are given in Section 5 for the univariate model-updating problem. A brief introduction of the Kriging model, and the description of the model updating method with Kriging are provided in the Appendix.

2. THE UNIVARIATE PROBLEM

Figure 1 shows a 3-dof mass-spring system. It consists of three masses and six springs. The stiffness k_2 is selected as the updating parameter in this case study, and the third eigenvalue of the system ω_3^2 is selected as the response to be used for updating k_2. Hence, for the Kriging model used as surrogate, k_2 is the input and ω_3^2 is the output. Other parameters in the system are determined as following.
$m_1 = 1.0\ kg, m_2 = 4.0\ kg, m_3 = 1.0\ kg$

$k_1 = k_3 = 0, k_4 = 1.9\ N/m$ *or* $k_4 = 2.0\ N/m, k_5 = 2.0\ N/m, k_6 = 1.0\ N/m$
The feasible range of parameter k_2 is chosen as [6.5,9.5].

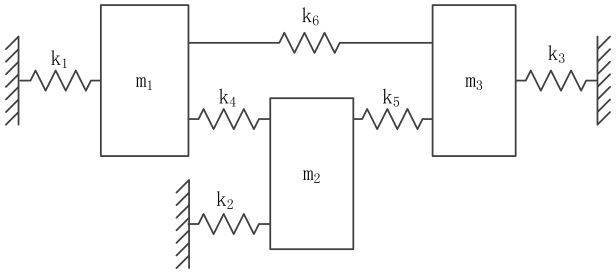

Figure 1 – 3-dof mass-spring system

Figure 2 presents the surfaces ω_2^2 and ω_3^2 with different k_2, k_4 values. The two surface are quite close to each other and connect at the point $(k_2, k_4) = (8, 2) N/m$, so that the response of the highest eigenvalue is represented by a curve $\omega_3^2(k_2)$ with a slope discontinuity when $k_4 = 2\ N/m$. The discontinuity then occurs at $k_2 = 8\ N/m$. This particular case is that of crossing eigenvalues and is chosen especially because it presents a challenging problem for the Kriging model and for the model updating. Any other value of k_4 results in veering of the eigenvalues. We select $k_4 = 1.9\ N/m$ and $k_4 = 2.0\ N/m$ as two cases for our particular study. The case of $k_4 = 1.9\ N/m$ is a more usual problem with a smooth response curve $\omega_3^2(k_2)$. The method described in the following two sections will be applied to these two problems and the results will be compared.

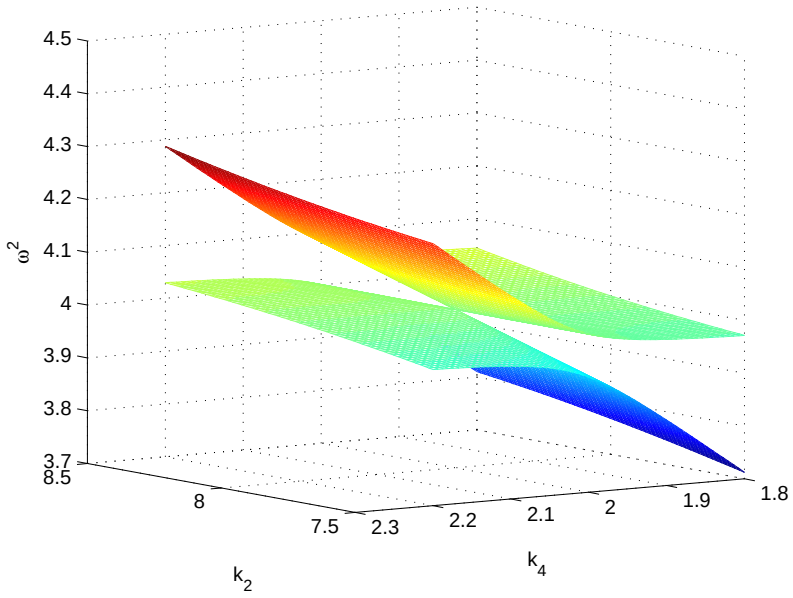

Figure 2 – ω_2^2, ω_3^2 *vs* k_2, k_4

3. BACKGROUND THEORY

A brief introduction of the background theory, including iterative model updating with a Kriging model, the principle of the kernel density estimation, and the bootstrap algorithm for density estimations, to be used in the model updating process will be presented in this section.

3.1. Model Updating with Kriging Predictor

If a Kriging predictor is constructed as a surrogate model for the mathematical (physical) model of the system. An iterative model updating method is to be used with the Kriging model to evaluate the possible parameter values.

This method is described by Haddad Khodaparast et al [5] who gave recursive equation for the updating parameter as (Eq. 25, [5])

$$\boldsymbol{\theta}_{l+1} = (\mathbf{H}^{\mathbf{T}}\mathbf{H} + \mathbf{D} + \mathbf{U} + \mathbf{V} - \mathbf{A} + \mathbf{W})^{-1}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_l} \{\mathbf{f}(\boldsymbol{\theta}) + \mathbf{H}^{\mathbf{T}}\boldsymbol{\mu} - \mathbf{H}^{\mathbf{T}}\boldsymbol{\Lambda}\boldsymbol{\rho} - \mathbf{g}(\boldsymbol{\theta}) + \mathbf{W}\boldsymbol{\theta}\}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_l} \tag{1}$$

This equation comes from the optimisation problem $\min_{\theta}(\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon})$, where $\boldsymbol{\varepsilon}$ is the error vector between measurement $\mathbf{z}_m$ and Kriging predictor output $\hat{\mathbf{z}}$. The explanation of the matrices in Equation 1 is provided in the Appendix. More details of the derivation of the recursive equation equation (1) can be found in [5].

3.2. Kernel Density Estimation

Kernel density estimation $f_X(x,\sigma) = \frac{1}{\sigma n\sqrt{2\pi}} \sum_{j=1}^{n} \exp\left(\frac{-(x-x^{(j)})^2}{2\sigma^2}\right)$ will be implemented for estimating the probability distribution of the updated parameters from a limited number of samples. As this kernel density estimator uses the normal distribution kernel, the smoothing parameter of the estimator, or the bandwidth h is equivalent to the standard deviation σ in the kernel density function. In the specific case of normal distributed target function, the bandwidth is given by $\sigma = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}}$, where $\hat{\sigma}$ is the standard deviation of the samples and n is the sample size. This estimation is optimal for estimating normal underlying densities. It should be noted that the σ in the kernel density estimation function is different from the $\hat{\sigma}$ in the bandwidth function. σ is the standard deviation of the kernel, and $\hat{\sigma}$ is the standard deviation of the samples to be estimated.

3.3. Bootstrap Algorithm for Density Estimation

The bootstrap is a re-sampling method for assigning measures of accuracy to sample estimates. It is the practice of estimating properties of an estimator by measuring those properties when sampling from an approximating distribution. Bootstrap algorithm can be utilised for estimating the variation of density estimates [6]. New estimates are constructed with re-sampled data and the variation of the estimates is obtained by by the repeated bootstrap re-sampling to determine re-constructed density estimates. The algorithm is described in the following steps:

If $\{x_i; i = 1,...,n\}$ is a set of data with probability density p. The bootstrap algorithm for this data is as follows [6]:

1. Form density estimate $\hat{p}$ from data $\{x_i; i = 1,...,n\}$.
2. Resample (uniformly) n values from $\{x_i; i = 1,...,n\}$, with replacement, obtaining $\{x_i^*; i = 1,...,n\}$ (bootstrap data). Note that in resampling with replacement our bootstrap dataset may contain the same x_i multiple times.
3. Form density estimate $\hat{p}^*$ from data $\{x_i^*; i = 1,...,n\}$.
4. Repeat steps 2&3 N times to obtain a family of bootstrap density estimates $\{\hat{p}_i^*; i = 1,...,N\}$.
5. The distribution of $\hat{p}_i^*$ about $\hat{p}$ mimics the distribution of $\hat{p}$ about p.

The bootstrap density estimates reproduce the variance of $\hat{p}$, and it is a useful tool for estimating the variance of a density estimator.

4. ESTIMATING OF THE PROBABILITY DISTRIBUTION OF THE UPDATED PARAMETERS

In this section, an approximate calculation method for determining the covariance matrix of the random variables is presented. The Jacobian matrix between the output and input of the Kriging model is derived and will be used for the covariance calculation by linearisation of the input-output relationship from the Kriging predictor. The Kriging prediction covariances will then be included in the model updating by re-sampling from the derived parameter covariances. Kernel density estimates are constructed with the re-sampled data and its confidence intervals are evaluated with the bootstrap algorithm.

4.1. Linear approximation using predicting variance from the Kriging model

In order to include the variance information given by the Kriging model in estimating parameters, a linear approximation will be used to transfer the Kriging output variance to parameter variance. We assume the parameter/output relationship to be linear in the local area around an updated parameter. The parameter covariance matrix can be determined from the outut prediction covariance matrix given by the Kriging model. A general form of a transformation between input vector $\boldsymbol{\theta}$ and output vector $\mathbf{Z}$ can be expressed as

$$\mathbf{Z} = \mathbf{J}\boldsymbol{\theta} \tag{2}$$

where $\mathbf{J}$ is the Jacobian matrix between X and Y. Matrix $\mathbf{J}$ can be presented as

$$\mathbf{J} = \begin{bmatrix} \frac{\partial z_1}{\partial \theta_1} & \frac{\partial z_1}{\partial \theta_2} & \cdots & \frac{\partial z_1}{\partial \theta_p} \\ \frac{\partial z_2}{\partial \theta_1} & \frac{\partial z_2}{\partial \theta_2} & \cdots & \frac{\partial z_2}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial z_n}{\partial \theta_1} & \frac{\partial z_n}{\partial \theta_2} & \cdots & \frac{\partial z_n}{\partial \theta_p} \end{bmatrix} \tag{3}$$

Then, the relationship between $\mathbf{Cov_Z}$, the covariance matrix of $\mathbf{Z}$, and $\mathbf{Cov_{\boldsymbol{\theta}}}$, the covariance matrix of $\boldsymbol{\theta}$, can be written as

$$\mathbf{Cov_Z} = \mathbf{J}\mathbf{Cov_{\boldsymbol{\theta}}}\mathbf{J}^T \tag{4}$$

By using this equation, we can transfer the variance between $\boldsymbol{\theta}$ and $\mathbf{Z}$.
Specifically, for the Kriging model used in this report, the Jacobian matrix for the model in Equation 15 can be written as

$$\mathbf{J} = \mathbf{J_I} + \mathbf{J_{II}} \tag{5}$$

where $\mathbf{J_I}$ is the Jacobian matrix of the regression part, $\beta_{0,i} + \mathbf{b}_i^T\boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\theta}^T\mathbf{B}_i\boldsymbol{\theta}$, of the Kriging model.

$$\mathbf{J_I} = \begin{bmatrix} \mathbf{b_1}^T + \boldsymbol{\theta}^T\mathbf{B}_1 \\ \mathbf{b_2}^T + \boldsymbol{\theta}^T\mathbf{B}_2 \\ \vdots \\ \mathbf{b_n}^T + \boldsymbol{\theta}^T\mathbf{B}_n \end{bmatrix}_{|\boldsymbol{\theta}=\boldsymbol{\theta}^*} \tag{6}$$

$\mathbf{J_{II}}$ is the Jacobian matrix for the other part, $\boldsymbol{\lambda}_i^T\mathbf{r}_i(\boldsymbol{\theta})$.

$$\mathbf{J_{II}} = \begin{bmatrix} \frac{\partial \boldsymbol{\lambda}_1^T\mathbf{r}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \frac{\partial \boldsymbol{\lambda}_2^T\mathbf{r}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial \boldsymbol{\lambda}_n^T\mathbf{r}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{bmatrix}_{|\boldsymbol{\theta}=\boldsymbol{\theta}^*} \tag{7}$$

where

$$\frac{\partial \boldsymbol{\lambda}_i^T\mathbf{r}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \boldsymbol{\lambda}_i^T\mathbf{r}_i(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \boldsymbol{\lambda}_i^T\mathbf{r}_i(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial \boldsymbol{\lambda}_i^T\mathbf{r}_i(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix} \tag{8}$$

$$\frac{\partial \boldsymbol{\lambda}_i^T\mathbf{r}_i(\boldsymbol{\theta})}{\partial \theta_t} = -\zeta_{t,i}\nu_i \cdot \lambda_i^T \begin{bmatrix} (-1)^s|\theta_t - \theta_t^{(1)}|^{\nu_i - 1}C_i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(1)}) \\ (-1)^s|\theta_t - \theta_t^{(2)}|^{\nu_i - 1}C_i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(2)}) \\ \vdots \\ (-1)^s|\theta_t - \theta_t^{(n_s)}|^{\nu_i - 1}C_i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(n_s)}) \end{bmatrix}, \quad s = \begin{cases} 0 & \theta_t \geq \theta_t^{(h)} \\ 1 & \theta_t < \theta_t^{(h)} \end{cases} \tag{9}$$

The equations above based on the differentiability of the Kriging model, which is related to the chosen of the correlation function.
For a single updated value d of the parameter from the Kriging model updating, with the calculated parameter variance σ_k, a new sample should be sampled from the distribution $N(d, \sigma_k)$. A kernel density estimate is constructed with the new samples of all the updated parameter values, and the confidence intervals of this density estimate is evaluated by bootstrapping. The whole process can be summarised as:

1. Deterministic model updating with the constructed Kriging model using equation (1).
2. Evaluate the variance of the updated parameters using equation (4).
3. Re-sample using a different dataset taken from the distribution of the updating parameters.
4. Evaluate the kernel density estimate of each of the new samples.
5. Bootstrap the new samples to evaluate the confidence interval of the density estimate.
6. Repeat from step 3.

5. NUMERICAL EXAMPLES

In this section, the numerical implementation of the model updating methodology will be presented using several numerical simulation case studies. The simulation will be carried out for the univariate problem of the mass-spring system. The procedure will be introduced in the following order.

- Multiple deterministic model updating and the updated parameter distribution determined using kernel density estimation. In this section, the uncertainty caused by the Gaussian process of the Kriging predictor will not be taken account.
- Determining the distribution of the input, or updating parameter (stiffness) from the known distribution on the output (natural frequency) by the approximate calculation using linearisation of the input-output relationship from the Kriging predictor. The uncertainty from the Kriging predictor will be taken account in the determination.
- Approximate calculation of output distribution by Monte Carlo simulation. The result of the simulation taking account and not taking account of the uncertainty from the Kriging model will both be presented.

For constructing the Kriging model, design points need to be sampled from the feasible range of k_2. As this is a univariate problem, the sampling of the Kriging predictor started with three points, $k_2 = 6.5, 8.0, 9.5$, two end points and the middle point in the feasible range. A sequential procedure was used for design point sampling with new design points added using the the maximal MSE criterion [2][7]. A quite coarse Kriging situation with only 4 design points was deliberately constructed for the two cases $k_4 = 1.9\ N/m$ and $k_4 = 2.0\ N/m$. Figure 3 shows the target functions given by mathematical models and the Kriging predictors. The second function is with slope discontinuity at $k_2 = 8\ N/m$ cannot be modelled accurately by Kriging with only 4 design points marked by circles in Figure 3 and the 95% confidence interval of the Kriging predictors shown as a shaded region. It is obvious that the Kriging model is less good in the case of the second target function. Surrogate models are needed as substitutes for expensive FE-models. The purpose of constructing a Kriging predictor here is to demonstrate a strategy for combining the Kriging predictor as a part of the a model updating procedure that fully takes account of the Kriging variance.

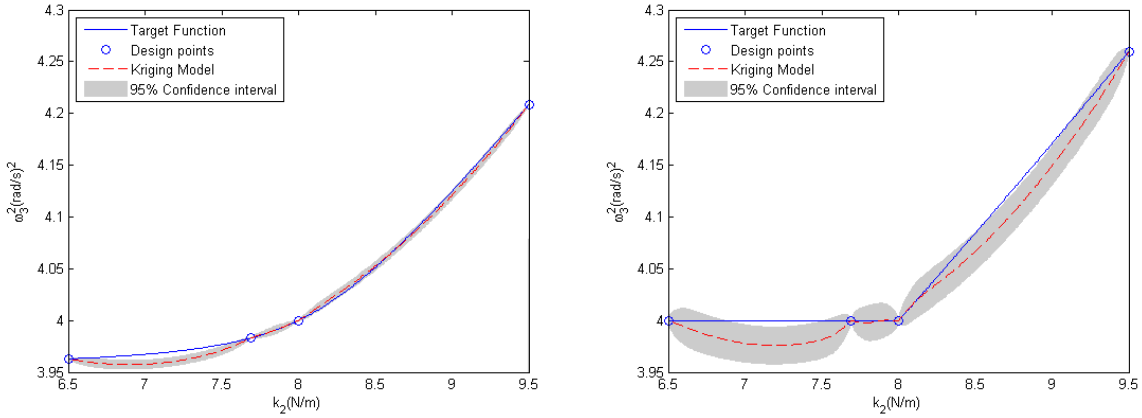

Figure 3 – The 4 point Kriging model of the univariate system. Left: $k_4 = 1.9\ N/m$. Right: $k_4 = 2.0\ N/m$.

The model updating process is based on the Kriging predictor. Two sets of synthetic observation samples are used for the comparison of the model updating results. The first sample sets consists of 10 samples, and the other one consists of 100 samples. The sample generating method is as follows.

1. Generating parameter k_2 from distribution $N(8.5, 0.5)$.
2. Calculating the corresponding $\omega_3^2(k_2)$ using the mathematical model.
3. Mixing the calculated ω_3^2 values with noise generated from $N(0, 0.001)$.

By using this procedure, the uncertainty in the measurement comes from two parts, the parameter uncertainty and measurement noise.

5.1. Estimation of Parameter Distribution using Deterministic Model Updating

Model updating was applied for each measurement sample, and the initial value of k_2 for updating was set as 8.5 for all the samples. After model updating, updated parameter mean values were obtained. The distribution of parameter was evaluated by kernel density estimation with the updated parameter mean values. The kernel function was chosen for normal distributions.

Updated results of two measurement sets for two cases are shown in Figure 4 and 5. Kernel density estimates for updated parameters were compared with the true parameter densities ($N(8.5, 0.5)$). The kernel density estimates differ from the true densities because of an insufficient number of Kriging training points and from each other because of the different numbers of samples used. But as the true parameter density would be unknown for a real application, this density estimation is the maximum information that can obtained based on the current method. Additionally, as the second function concerned in this case study contains a horizontal line where the eigenvalue is insensitive to the parameter, the model updating results in this region cannot be expected to be accurate. This problem also influences the accuracy of the parameter kernel density estimate. The parameter updating of strictly non-monotonic functions would be a topic to be studied in the future.

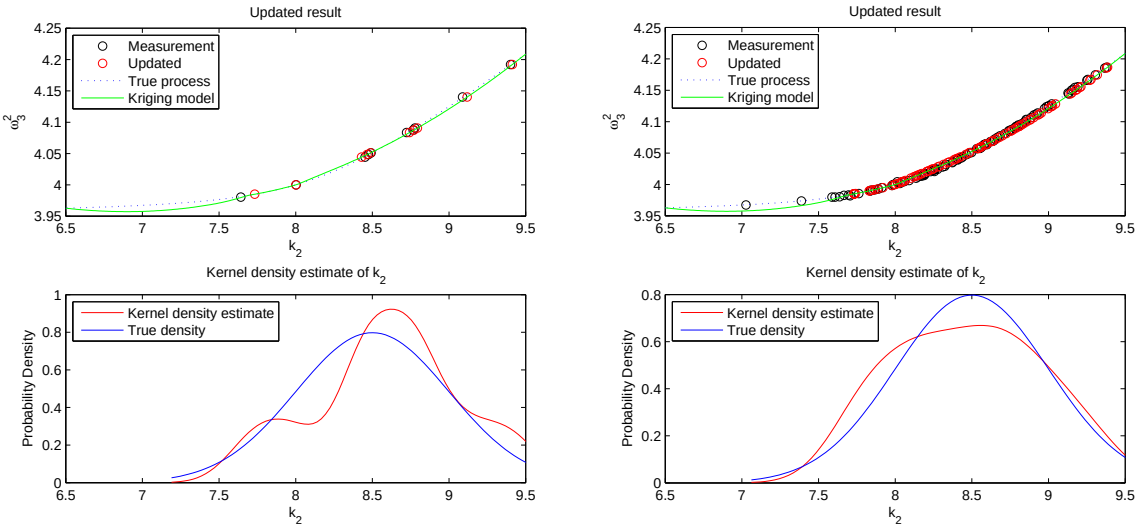

Figure 4 – Model updating results of 4 point Kriging model, $k_4 = 1.9\ N/m$. Left: 10 samples. Right: 100 samples.

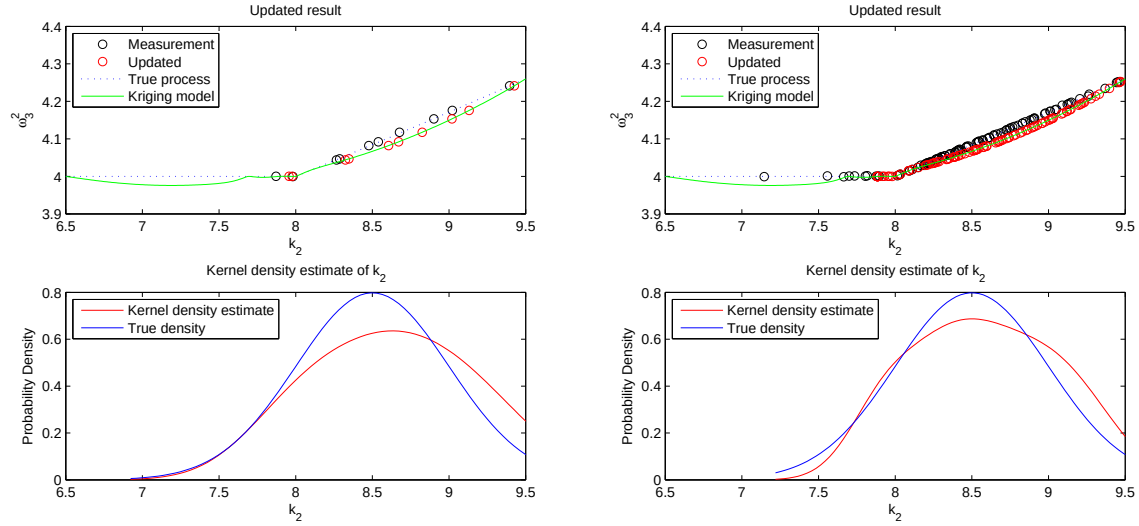

Figure 5 – Model updating results of 4 point Kriging model, $k_4 = 2.0\ N/m$. Left: 10 samples. Right: 100 samples.

5.2. Inclusion of Kriging Variance Information

The principle of the model updating is to find the values of parameters to minimise the difference between the response calculated with parameters and the measured response. In the ideal situation, the updating process will lead to a parameter value that reproduces the measured response exactly . If the response function is monotonic and the surrogate model reflects the true model perfectly, the updating process will find the accurate value for the

updating parameter. However, the Kriging model only holds for perfect fitting at its design points. At other points, the response calculated by the Kriging model is usually slightly different from the true model. As a result of this small difference, the updated parameter values found with the Kriging model will slightly biased from the correct value. The model updating based on the mean value of the Kriging model gives only one possible result in the confidence interval of the Kriging model. There will be other valid points within the density of the Kriging outputs that should be considered.

5.3. Updating using Kriging Variance Data and Bootstrapping

Figures 6 and 7 show the comparison of results from two different sample sizes for the two cases concerned. The true distribution is denoted by the blue curve. The red curves are kernel density estimates obtained by sampling from parameter densities obtained by inversion of equation (4). The blue dashed lines are 95% confidence intervals obtained by bootstrapping. It can be seen that the confidence intervals in all cases enclose the red curves. Also the general shape of the red curves and boundaries of the confidence intervals are similar. As the sample size increases, the variation of the re-constructed density estimates reduces and the confidence intervals become narrower. The crossing-modes curve with the greater Kriging prediction variance also creates greater variance in the parameter densities. The results for the $k_4 = 2.0\ N/m$ case with a 12 design point model is shown in Figure 8. As the Kriging prediction variance gets smaller with more design points, the variation of the result also becomes smaller compared to the 4 design-point case.

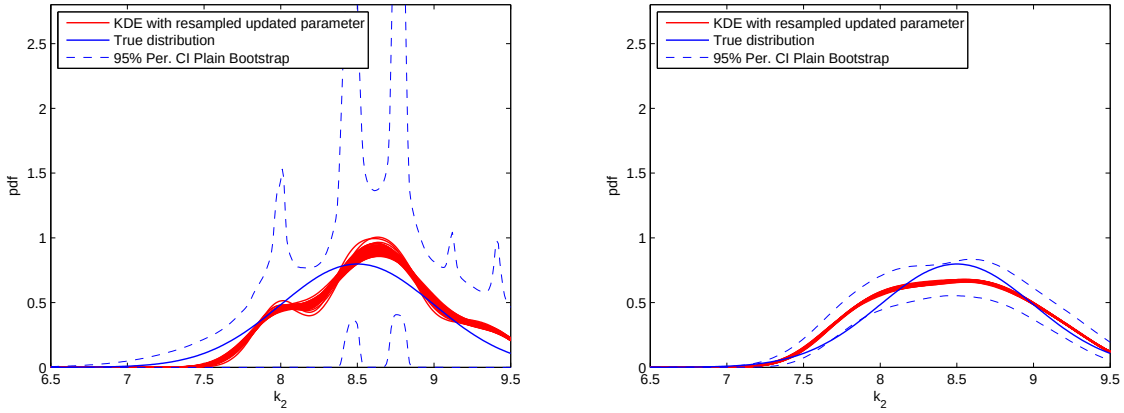

Figure 6 – Bootstrapping results of the 4 point Kriging model, $k_4 = 1.9\ N/m$. Left: 10 samples. Right: 100 samples.

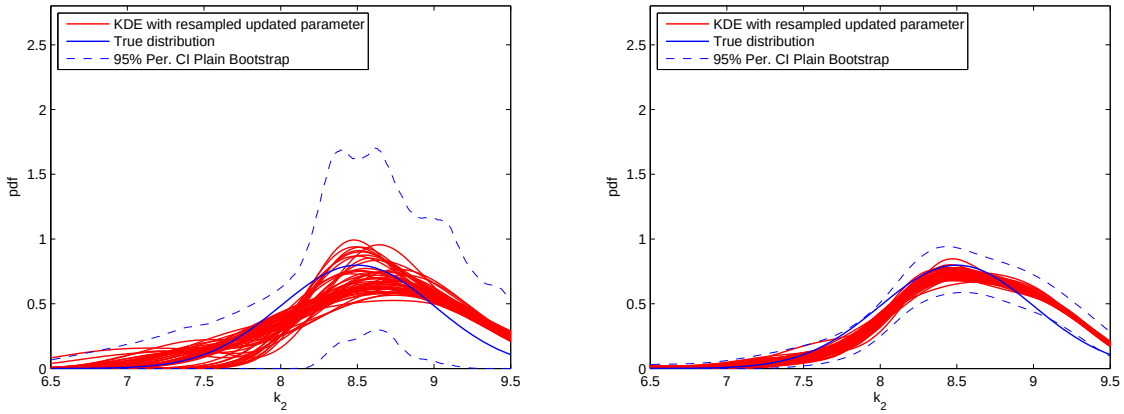

Figure 7 – Bootstrapping results of the 4 point Kriging model, $k_4 = 2.0\ N/m$. Left: 10 samples. Right: 100 samples.

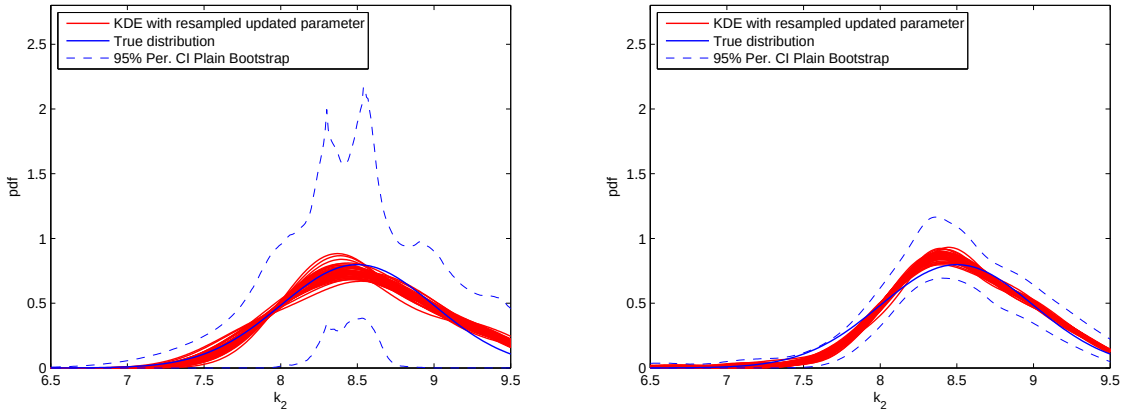

Figure 8 – Bootstrapping results of the 12 point Kriging model, $k_4 = 2.0\ N/m$. Left: 10 samples. Right: 100 samples.

6. CONCLUSION

This paper addresses the use of Kriging model in model updating and presents a method to incorporate the effect of Kriging prediction variance in the estimation of model-updating parameters. Numerical implementations for a univariate problem are described and the results obtained show the effect of Kriging prediction variance on the model updating result. It shows that the Kriging prediction variance does produce an evident influence on model updating results based on a Kriging model. Therefore for obtaining a more precise result, the prediction variance of Kriging model should be considered in the model updating procedure and the method provided in the article could be used to estimate the effect of Kriging prediction variance.

APPENDIX

Kriging predictor

A Universal Kriging predictor can be presented as

$$Z(\boldsymbol{\theta}) = \sum_{j=1}^{k} \beta_j f_j(\boldsymbol{\theta}) + \varepsilon(\boldsymbol{\theta}) \tag{10}$$

The random function $Z(\boldsymbol{\theta})$ includes a regression model $\sum_{j=1}^{k} \beta_j f_j(\boldsymbol{\theta})$, and a random process $\varepsilon(\boldsymbol{\theta})$. $\varepsilon(\boldsymbol{\theta})$ is assumed to have mean zero and covariance $Cov(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = \sigma^2 C(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ between $\varepsilon(\boldsymbol{\theta})$ and $\varepsilon(\boldsymbol{\vartheta})$, where σ^2 is the process variance and $C(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ is the correlation.

In this report, a second order polynomial is used for the regression part in the Kriging model. Hence, the Kriging model of the i-th output is expressed as

$$\hat{z}_i = \beta_{0,i} + \mathbf{b}_i^T \boldsymbol{\theta} + \frac{1}{2} \boldsymbol{\theta}^T \mathbf{B}_i \boldsymbol{\theta} + \varepsilon_i(\boldsymbol{\theta}) \tag{11}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]^T$ is the uncertain system parameter vector. $\mathbf{b}_i^T = [\beta_{1,i}, \beta_{2,i}, \cdots, \beta_{p,i}]^T$,

$\mathbf{B}_i = \begin{bmatrix} 2\beta_{11,i} & \beta_{12,i} & \cdots & \beta_{1p,i} \\ & 2\beta_{22,i} & \cdots & \beta_{2p,i} \\ & & \ddots & \vdots \\ sym. & & & 2\beta_{pp,i} \end{bmatrix}$, and the covariance of $\varepsilon_i(\boldsymbol{\theta})$ is

$$cov(\varepsilon_i(\boldsymbol{\theta}), \varepsilon_i(\boldsymbol{\vartheta})) = \sigma_i^2 C_i(\boldsymbol{\theta}, \boldsymbol{\vartheta}) \tag{12}$$

$$C_i(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = \prod_{j=1}^{p} C_{j,i}(\theta_j, \vartheta_j) \tag{13}$$

$$C_{j,i}(\theta_j, \vartheta_j) = exp(-\zeta_{j,i} |\theta_j - \vartheta_j|^{\nu_i}), 1 \leq \nu_i \leq 2 \tag{14}$$

where $\zeta_{j,i}$ and ν_i are the parameters of the correlation function at the i-th output.

The mean Kriging predictor may be expressed as

$$\hat{z}_i = \beta_{0,i} + \mathbf{b}_i^T \boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\theta}^T \mathbf{B}_i \boldsymbol{\theta} + \boldsymbol{\lambda}_i^T \mathbf{r}_i(\boldsymbol{\theta}) \tag{15}$$

where $\mathbf{r}_i(\boldsymbol{\theta}) = [C_i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(1)}), C_i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(2)}), \cdots, C_i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(n_s)})]^T$, $\boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(n_s)}$ are the design points, and $\boldsymbol{\lambda}_i = \mathbf{R}_i^{-1}(\mathbf{Z}_{:,i} - \boldsymbol{\Xi}[\beta])$

The training of the Kriging predictor starts with a few initial training points. Ghoreyshi et al [2] used Latin hypercube sampling (LHS) to generate initial points and a central composite design (CCD) was used in [5]. In this study, a one dimensional system is used and the initial points are chosen as three points which are the two boundaries points and the middle point of the region. Sacks et al [1] introduced several design criteria including Integrated Mean Squared Error (IMSE), Maximum Mean Squared Error (MMSE), and Entropy. The MMSE criterion will be used in this study. The maximum point of mean squared error (MSE) of Kriging predictor will be used to determine the location of the new point in iteration.

For a Kriging predictor expressed as equation (10). The MSE for this predictor can be expressed as

$$MSE(\hat{z}_i(\boldsymbol{\theta})) = \sigma_i^2 [1 - \begin{pmatrix} \mathbf{f}^T(\boldsymbol{\theta}) & \mathbf{r}^T(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} 0 & \mathbf{F}^T \\ \mathbf{F} & \mathbf{R} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{f}(\boldsymbol{\theta}) \\ \mathbf{r}(\boldsymbol{\theta}) \end{pmatrix} \tag{16}$$

where $\hat{z}_i(\boldsymbol{\theta})$ is the linear predictor of $z_i(\boldsymbol{\theta})$ at an untried $\boldsymbol{\theta}$. $\mathbf{f}(\boldsymbol{\theta}) = [f_1(\boldsymbol{\theta}), \cdots, f_k(\boldsymbol{\theta})]^T$ is the vector of k functions in the regression, $\mathbf{F} = \begin{pmatrix} \mathbf{f}^T(\boldsymbol{\theta}^{(1)}) \\ \vdots \\ \mathbf{f}^T(\boldsymbol{\theta}^{(n_s)}) \end{pmatrix}$ is the $n_s \times k$ expanded design matrix, $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(j)}), 1 \le i \le n_s; 1 \le j \le n_s$ is the $n_s \times n_s$ matrix of stochastic-process correlations between the design points, and $\mathbf{r}(\boldsymbol{\theta}) = [R(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}), \cdots, R(\boldsymbol{\theta}^{(n_s)}, \boldsymbol{\theta})]^T$ is the vector of correlations between design points and an untried $\boldsymbol{\theta}$.

The goal of the MMSE criterion is to minimise

$$\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} MSE[(\hat{z}_i(\boldsymbol{\theta})] \tag{17}$$

In each training step, the MSE of the predictor is calculated and a maximum MSE point will be added to training points in the next training step. Thus, the MSE of predictor will be reduced step by step with the increasing of the number of training points. The training iteration stopped when the maximum MSE falls under a chosen threshold.

Model updating recursive equation

In the model updating recursive equation

$$\boldsymbol{\theta}_{l+1} = (\mathbf{H^T H} + \mathbf{D} + \mathbf{U} + \mathbf{V} - \mathbf{A} + \mathbf{W})^{-1}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_l} \{\mathbf{f}(\boldsymbol{\theta}) + \mathbf{H^T}\boldsymbol{\mu} - \mathbf{H^T}\boldsymbol{\Lambda}\boldsymbol{\rho} - \mathbf{g}(\boldsymbol{\theta}) + \mathbf{W}\boldsymbol{\theta}\}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_l}$$

the process matrices $\mathbf{H}, \mathbf{D}, \mathbf{U}, \mathbf{V}, \mathbf{A}, \mathbf{f}, \mathbf{g}, \mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Lambda}$ are

$$\mathbf{H}(\boldsymbol{\theta}) = [H_{ij}]_{n \times p}; \qquad H_{ij} = \beta_{j,i} + \theta_j \beta_{jj,i} + \frac{1}{2} \sum_{k=1, k \neq j}^{p} \beta_{kj,i\theta_k}$$

n – the number of the Kriging outputs, p – the number of Kriging inputs, n_s – the number of the Kriging design points

$$\mathbf{D}(\boldsymbol{\theta}) = [D_{ij}]_{p \times p}; \quad D_{ij} = \frac{1}{2} \sum_{k=1}^{p} \sum_{l=1}^{n} \left(H_{lj} \frac{\partial H_{lj}}{\partial \theta_i} + H_{lk} \frac{\partial H_{lj}}{\partial \theta_i} \right) \theta_k$$

$$\mathbf{U}(\boldsymbol{\theta}) = [U_{ij}]_{p \times p}; \quad U_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n_s} \lambda_{l,k} H_{kj} \frac{\partial r_{l,k}(\theta)}{\partial \theta_i}$$

$$\mathbf{V}(\boldsymbol{\theta}) = [V_{ij}]_{p \times p}; \quad V_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n_s} \lambda_{l,k} \frac{\partial H_{kj}}{\partial \theta_i} r_{l,k}(\theta)$$

$$\mathbf{A} = [A_{ij}]_{p \times p}; \quad A_{ij} = \sum_{k=1}^{n} \mu_k \frac{\partial H_{ij}}{\partial \theta_i}; \quad \frac{\partial H_{ij}}{\partial \theta_i} = \frac{1}{2} B_{ij,k}$$

$$\mathbf{f}(\boldsymbol{\theta}) = \{f_i(\theta)\}_{p \times 1}; \quad f_i(\theta) = \sum_{j=1}^{n} \sum_{k=1}^{n_s} \lambda_{k,j} \mu_j \frac{\partial r_{k,j}(\theta)}{\partial \theta_i}$$

$$\mathbf{g}(\boldsymbol{\theta}) = \{g_i(\theta)\}_{p\times 1}; \quad g_i(\theta) = \frac{1}{2}\sum_{l=1}^{n}\sum_{k=1}^{n_s}\sum_{j=1}^{n_s}\lambda_{j,l}\lambda_{k,l}\left(r_{k,l}(\theta)\frac{\partial r_{j,l}(\theta)}{\partial\theta_i} + r_{j,l}(\theta)\frac{\partial r_{k,l}(\theta)}{\partial\theta_i}\right)$$

$$\boldsymbol{\mu} = \mathbf{z}_m - \boldsymbol{\beta}_0$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{1,1} & \lambda_{2,1} & \cdots & \lambda_{n_s,1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & \lambda_{1,2} & \cdots & \lambda_{n_s,2} & 0 & \cdots & 0 \\ & & & & & & & & \cdots & \\ & & & & & & & & \cdots & \lambda_{n_s,n} \end{bmatrix}$$

$$\boldsymbol{\lambda}_i = (\lambda_{i,1}, \lambda_{i,2}, \cdots, \lambda_{i,n})^T = \mathbf{R}_i^{-1}(\mathbf{Z}_i - \mathbf{F}\boldsymbol{\beta})$$

Matrix $\mathbf{W}$ is chosen to make $(\mathbf{W^T W} + \mathbf{D} + \mathbf{U} + \mathbf{V} - \mathbf{A} + \mathbf{W})$ invertible. It may be chosen in the form $\mathbf{W} = \lambda\mathbf{I}$. $B_{ij,k}$ is the component of matrix $\mathbf{B}_k$ which is a parameter matrix in the Kriging predictor expression equation (11).

REFERENCES

[1] Sacks, J., Welch, W., Mitchell, T., Wynn, H., "Design and Analysis of Computer Experiments", Statistical Science, 4(4), 1989, pp. 409–423.

[2] Ghoreyshi, M., Badcock, K.J., Woodgate, M.A., "Accelerating the Numerical Generation of Aerodynamic Models for Flight Simulation", Journal of Aircraft, 46(3), 2009, pp. 972–980.

[3] Kleijnen, J.P.C., Beers, W.C.M.v., "Application-Driven Sequential Designs for Simulation Experiments: Kriging Metamodelling", The Journal of the Operational Research Society, 55(8), 2004, pp. 876–883.

[4] Jones, D., "A Taxonomy of Global Optimization Methods Based on Response Surfaces", Journal of Global Optimization, 21(4), 2001, pp. 345–383.

[5] Khodaparast, H.H., Mottershead, J.E., Badcock, K.J., "Interval model updating with irreducible uncertainty using the Kriging predictor", Mechanical Systems and Signal Processing, 25(4), 2011, pp. 1204–1206.

[6] Narsky, I., Porter, F.C., Density Estimation, Wiley-VCH Verlag GmbH & Co. KGaA, 2013, pp. 89–120.

[7] Kleijnen, J.P.C., "Kriging metamodeling in simulation: A review", European Journal of Operational Research, 192(3), 2009, pp. 707–716.

[8] Silverman, B.W., Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986.